DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal

Zheng, Shuyu

2021-07-02

Zheng , S , Aldahdooh , J , Shadbahr , T , Wang , Y , Aldahdooh , D , Bao , J , Wang , W & Tang , J 2021 , ' DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal ' , Nucleic Acids Research , vol. 49 , no. W1 , pp. W174-W184 . https://doi.org/10.1093/nar/gkab4

http://hdl.handle.net/10138/333475
https://doi.org/10.1093/nar/gkab438

W174–W184 *Nucleic Acids Research, 2021, Vol. 49, Web Server issue* *Published online 1 June 2021*
https://doi.org/10.1093/nar/gkab438

DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal

Shuyu Zheng[1], Jehad Aldahdooh[1], Tolou Shadbahr[1], Yinyin Wang[1], Dalal Aldahdooh[1], Jie Bao[1,2], Wenyu Wang[1] and Jing Tang[1,*]

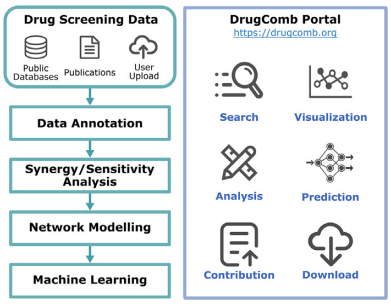
[1]Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki FI-00290, Finland and [2]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki FI-00290, Finland

Received March 24, 2021; Revised April 18, 2021; Editorial Decision April 28, 2021; Accepted May 06, 2021

ABSTRACT

Combinatorial therapies that target multiple pathways have shown great promises for treating complex diseases. DrugComb (https://drugcomb.org/) is a web-based portal for the deposition and analysis of drug combination screening datasets. Since its first release, DrugComb has received continuous updates on the coverage of data resources, as well as on the functionality of the web server to improve the analysis, visualization and interpretation of drug combination screens. Here, we report significant updates of DrugComb, including: (i) manual curation and harmonization of more comprehensive drug combination and monotherapy screening data, not only for cancers but also for other diseases such as malaria and COVID-19; (ii) enhanced algorithms for assessing the sensitivity and synergy of drug combinations; (iii) network modelling tools to visualize the mechanisms of action of drugs or drug combinations for a given cancer sample and (iv) state-of-the-art machine learning models to predict drug combination sensitivity and synergy. These improvements have been provided with more user-friendly graphical interface and faster database infrastructure, which make DrugComb the most comprehensive web-based resources for the study of drug sensitivities for multiple diseases.

GRAPHICAL ABSTRACT



INTRODUCTION

Despite the scientific advances in the understanding of complex diseases such as cancer, there remains a major gap between the vast knowledge of molecular biology and effective treatments. Next generation sequencing has revealed intrinsic heterogeneity across cancer samples, which partly explain why patients respond differently to the same therapy (1). For the patients that lack common oncogenic drivers, multi-targeted drug combinations are urgently needed, which shall block the emergence of drug resistance and therefore achieve sustainable efficacy (2). To facilitate the discovery of drug combination therapies, high-throughput drug screening techniques have been developed to allow for a large scale of drug combinations to be tested for their sensitivity (percentage inhibition of cell growth) and synergy (degree of interaction) in-vitro (3). Furthermore, patient-derived cancer cell cultures and xenograft models have been developed, which make the drug discovery closer to the actual patients (4–6).

With the increasing amount of drug sensitivity screening data, the challenge of translating them into actual drug discovery remains, as recent studies showed that most of clinically approved drug combinations work independently (7), that the efficacy and synergy observed in a pre-clinical setting may not be translated into a clinical trial (8,9). The challenge of utilizing the results from drug combination screens largely resides from un-harmonized metrics for syn-

*To whom correspondence should be addressed. Tel: +35 845 868 9708; Email: jing.tang@helsinki.fi

ergy and sensitivity that are derived from different mathematical models, which are often incompatible for the same datasets (10). Another limitation is the lack of standardization of drug combination experimental design and the insufficient level of data curation and deposition to publicly available databases (11). Furthermore, the drug combination data has not been harmonized with single drug screening data, partially due to a lack of computational tools to enable a systematic comparison of drug combination efficacy against single drug efficacy (12).

To initialize the efforts for curating drug combination datasets, and to facilitate a community-driven standardization of evaluation of the degree of synergy and sensitivity of drug combinations, we have provided DrugComb as the very first data portal to harbour the manually curated datasets as well as the web server to analyse them (13). The original version of DrugComb consists of four major high-throughput studies, which served as a reference dataset for developing machine learning algorithms to predict drug combination sensitivity and synergy (14). Different from other recent databases including DrugCombDB (15) and SynergxDB (16), DrugComb is a unique resource as it is a compendium of database and web application, not only for depositing deeply curated public datasets but also for the analysis and annotation of user-uploaded data. Furthermore, DrugComb provides detailed visualization of drug combination sensitivity and synergy, which shall greatly facilitate the understanding of drug interactions at specific dose levels. The data from DrugComb has been used to develop machine learning models for drug combination prediction (17,18), and synthetic lethality knowledge graph (14). The analysis tools provided by DrugComb have also helped to explore the mechanism of replication stress response in colorectal cancer stem cells (19).

With the development of high-throughput screening techniques, the number of data points for drug combinations has been greatly increased. For example, the recent Dream Challenge on drug combination prediction has provided more than 20k drug combinations in cancer cell lines (20). Furthermore, drug combination screening has been extended to other disease models such as malaria and Ebola (21). More recently, drug combination screening studies on COVID19 have been conducted, providing important clues for the treatment of the ongoing pandemic (22). In the new version of DrugComb, we aim to expand our manual curation from cancer to other diseases to improve the data coverage. On the other hand, drug combinations need to be harmonized with the monotherapy drug screening data, since these treatment options shall be evaluated using the same endpoint metric (such as progression free survival and overall survival) in clinical trials. Therefore, we aim to harmonize the drug combination with monotherapy drug screening, by providing informatics tools to evaluate their overall sensitivity in a more systematic manner. For this reason, in the new version of DrugComb, we do not limit ourselves for curating drug combination data, but rather we included monotherapy drug sensitivity screening data as well. More importantly, we provide a robust metric to enable a direct comparison of drug combinations and single drugs, as monotherapy drug screening can be considered as a subset of drug combination experiments. The new data harmonization framework thus allows a more systematic evaluation of a drug combination in comparison to a single drug. In addition, we implement several new modules for the analysis of these datasets, including the integration of drug targets and gene expressions of neighbouring proteins in a signalling network, such that the mechanisms of action of a drug or a drug combination can be annotated systematically in a specific cellular context. We also provide a baseline model based on CatBoost to predict the sensitivity and synergy of drug combinations, with which the machine learning community may develop novel algorithms to improve our understanding of drug responses in cancer cells. Taken together, the new version of DrugComb features an enhanced web portal to make drug screening data more interpretable and reusable for various applications such as machine learning, network modelling and experimental validation.

RESULTS

Overview of the DrugComb portal

DrugComb portal consists of two major components including a database for harbouring the most recent drug screening datasets as well as a web server to analyse and visualize these datasets or user-uploaded datasets for the degree of sensitivity and synergy. For retrieving the database, users can query by drug names, cell line names as well as study names. For utilizing the web server to analyse user-uploaded datasets, users need to import the data according to the format of an example file, and the results will be shown as both tabular and image displays, which are also downloadable. When users plan a drug combination experiment, they may utilize the web server to predict the sensitivity and synergy and utilize such information to guide the selection of drugs. The drug targets as well as the gene expressions of the signalling pathways for a given cancer cell line can be also annotated as a network model. In the following, we describe how we have improved the coverage of the database as well as the data analysis modules of the web server with a range of algorithms, and the new implementation techniques to accelerate data curation and harmonization efficiency (Figure 1).

Data sources

The initial version of DrugComb consists of four drug combination screening studies, covering 437 923 drug combination experiments. We have curated much more drug combination experiments for cancer cell lines. Furthermore, we have incorporated monotherapy drug screening datasets and considered them as a subset of a drug combination experiment, where the other drug is absent. We have also included the drug screening results from patient-derived cancer samples in haematological malignancies (5). In addition to multiple cancer types, we have extended the curation efforts to other diseases such as Ebola, malaria and COVID-19. The manual curation is under high level of quality control, that only those studies that reported the raw dose-response results will be considered, and thus the studies that reported only summary-level results including IC50, AUC (area under the dose response curves) or synergy scores (e.g. combination index) are excluded. We uti-

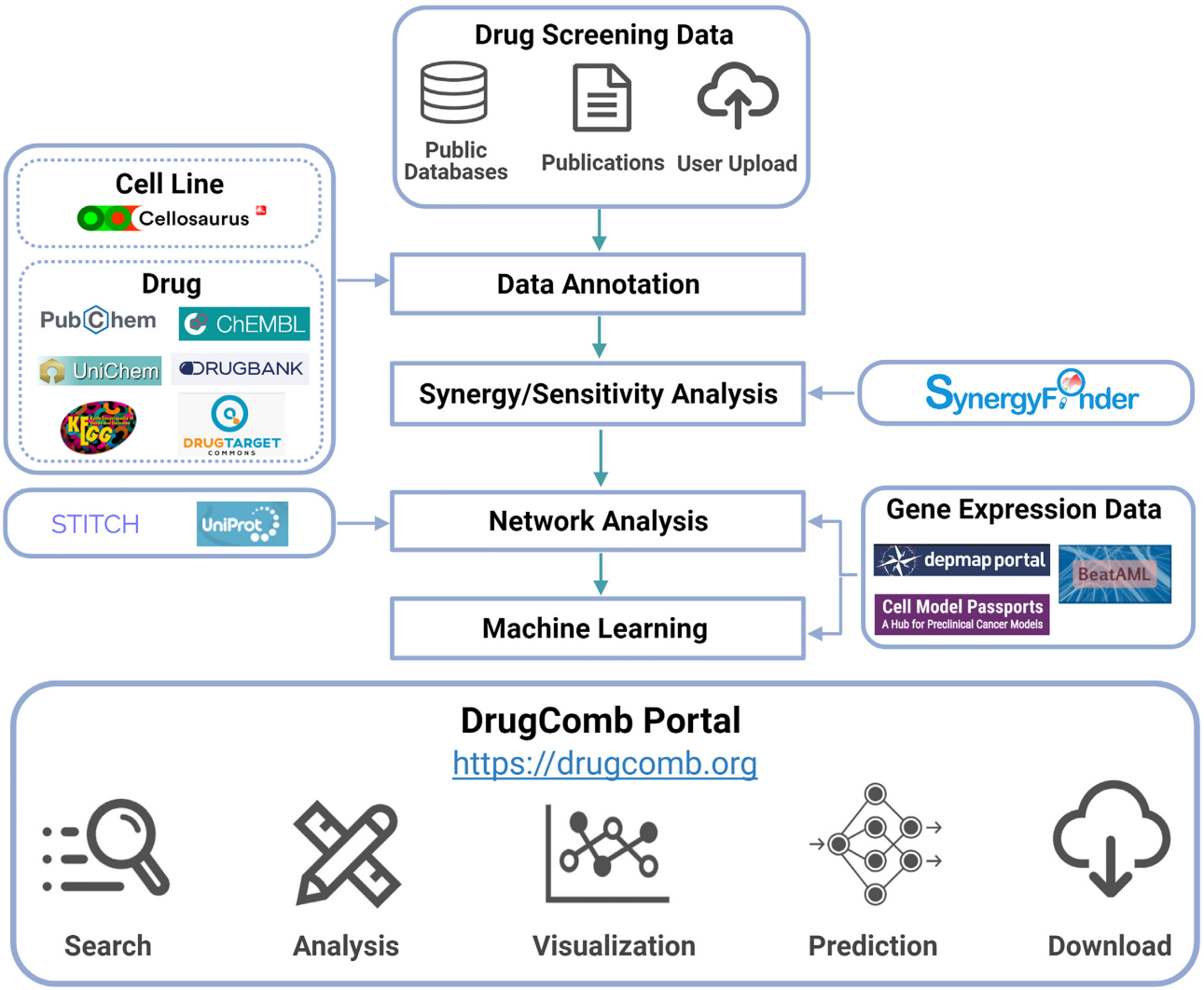


Figure 1. A schematic overview of the DrugComb database and web server pipeline. Drug combination and monotherapy drug screening datasets are curated from public databases, publications or user-upload. After quality control and pre-processing, the cell information is retrieved from Cellosaurus (23), while the drug information is retrieved from multiple databases including PubChem (24), ChEMBL (25), UniChem (26), DrugBank (27), KEGG (28) and DrugTargetCommons (29). The degree of synergy in drug combinations, as well as the sensitivity of drug combinations and single drugs are determined using the SynergyFinder R package (3). For inferring the mechanisms of action of drugs or drug combinations, their targets as well as interacting proteins are visualized in a signalling network, retrieved from STITCH (30) and UniProt (31). Furthermore, the gene expressions of these proteins in the given cancer cells are obtained from DepMap (32) and Cell Model Passports (33), and from BeatAML where the cancer samples were derived from AML (Acute Myeloid Leukaemia) patients (5). Machine learning algorithms utilize chemical structural and gene expression features to predict drug combination synergy and sensitivity. The DrugComb portal enables the query and download of curated raw datasets and analysis results, as well as the contribution of new datasets.

lized SynergyFinder (3) to determine the synergy scores directly from the raw dose-response data and compared them with those reported in the original publications. Only the datasets that have a correlation higher than 0.6 will be included. Furthermore, dose-response matrices containing abnormal response values, for example percentage inhibition of cell growth less than −200% or larger than 200%, were marked as poor-quality data points for which the data analysis results were not shown in the web interface. We have also standardized the metadata about experimental protocols of these studies so that their differences can be evaluated more systematically. The annotation of the bioassay protocols is based on the BAO (Bioassay annotation ontology) (34), that is commonly adopted for major chemical biology databases including ChEMBL (25), PubChem (24) and DrugTargetCommons (29). For the drugs and cell lines we provided the cross-database references such that their pharmacological and clinical information can be easily accessed (Figure 2A and B). As of March 2021, 751 498 drug combinations, 717 684 single drug screenings from 37 studies are deposited in DrugComb, corresponding to 21 621 279 unique data points spanning 2320 cell lines including 225 cancer types and three infectious diseases. Supplementary Table S1 shows the summary of the data points from the individual studies that are curated and harmonized in DrugComb.

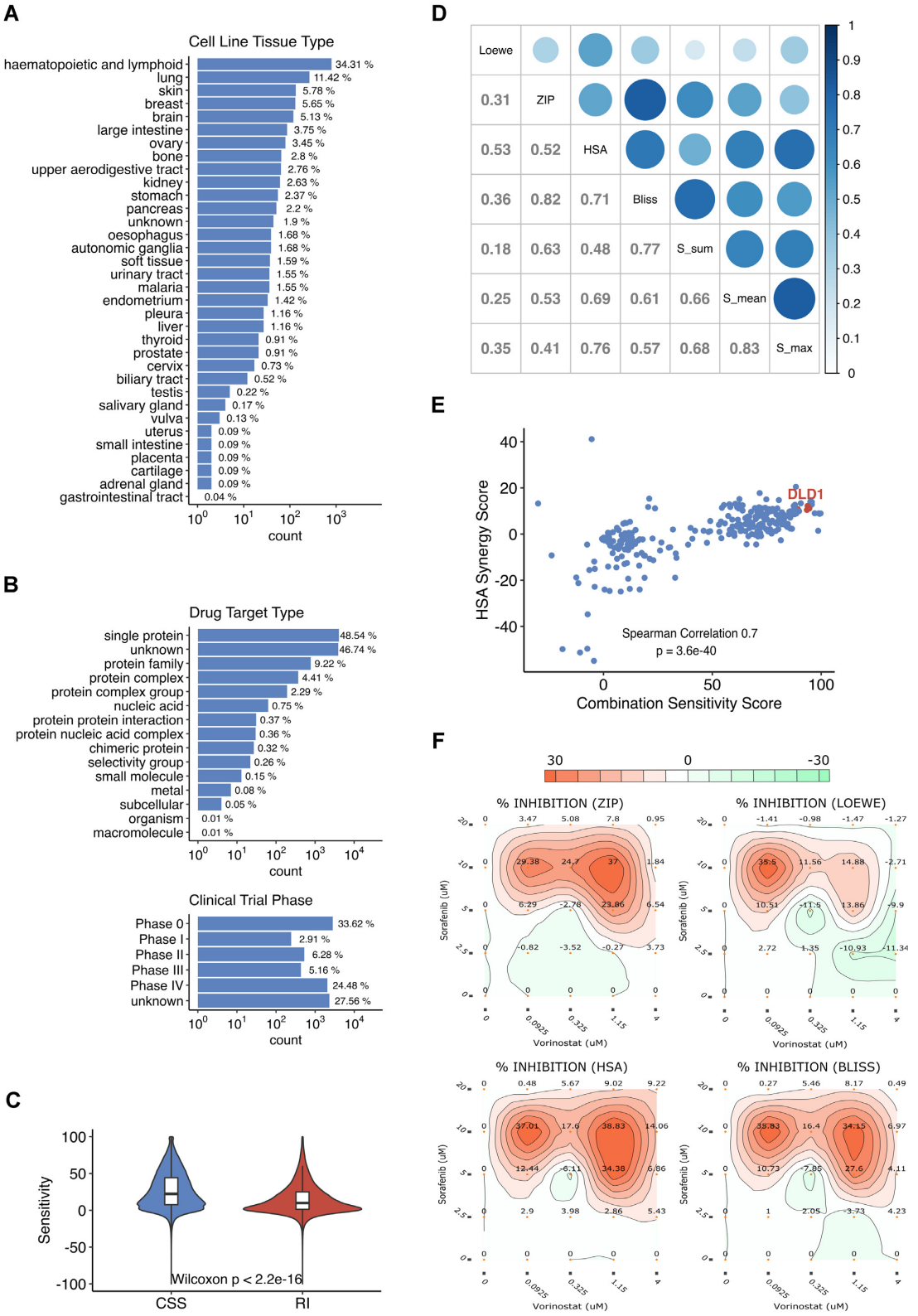


Figure 2. Overview of DrugComb data statistics. (**A**, **B**) Classifications of cell lines ($n = 2320$) and drugs ($n = 8397$). (**C**) The CSS score for drug combinations is higher than the RI score for monotherapy drugs, suggesting the general rationale for drug combination studies. (**D**) The correlations of synergy scores. (**E**) An example of SS plot for vorinostat and sorafenib combination across 128 cell lines. DLD-1 is a colon cancer cell line, which has shown strong synergy and sensitivity to the combination (38). (**F**) The synergy landscape over the dose-response matrix of vorinostat and sorafenib in DLD-1.

Algorithms for assessing sensitivity and synergy

DrugComb utilizes the SynergyFinder R package to analyse drug combination sensitivity and synergy. The single drug sensitivity is characterized as a dose-response curve with its IC50 and RI (relative inhibition) values. RI is the normalized area under the $\log_{10}$-transformed dose-response curves, which has shown enhanced robustness to characterize drug sensitivity (35). Moreover, RI can be interpreted as percentage inhibition, summarizing the overall drug inhibition effects relative to positive controls. With the RI metric, drug responses of different concentration ranges can be compared, in contrast to IC50 or EC50, which are usually a relative term depending on the tested concentration ranges.

For drug combination sensitivity, we provide a metric called CSS (Combination Sensitivity Score), that is based on the normalized area under the $\log_{10}$-transformed of the combination dose-response curve when one of the two drugs is fixed at its IC50 concentration (12). CSS and RI use the same principle to characterize the overall drug response efficacy, such that their values can be directly compared (Figure 2C). For evaluating the drug synergy, we implement four major mathematical models including Bliss, Loewe, HSA and ZIP (36) and provide the visualization of these scores in the dose-response matrices. Furthermore, we provide a synergy score called S score that is derived from the difference between CSS and RI scores of the combination and single drugs respectively (12). Drug combinations with synergy scores of zero are considered additive, while a positive synergy score suggests synergy, and a negative score suggests antagonism. The five synergy scores are based on different mathematical assumptions such that they do not necessarily match with each other (Figure 2D). For example, the Bliss model assumes probabilistic independence when drugs are non-interactive while the Loewe model assumes that the efficacy of non-synergistic drug combinations is identical to that of a drug combined with itself. The ZIP model, on the other hand, can be considered as an Ensembl model as it combines the assumptions of Bliss and Loewe (36). In actual clinical trials, approval of a drug combination often is based on the HSA model that simply shows that the drug combination improves patient survival compared to monotherapies. To insure the clinical translation of drug combinations, we encourage the use of all the major synergy scoring metrics, such that the top hits that pass the threshold of all of them can be prioritized (37). On the other hand, there have been biases by focusing solely on the synergy, while the sensitivity of a drug combination might be understudied. It is likely that a drug combination produces strong synergy while their overall efficacy is not achieving therapeutic relevance. Therefore, we provide an SS (Synergy-Sensitivity) plot to ensure that both of these two scores can be evenly weighted when interpreting the relevance of a drug combination (Figure 2E, Supplementary Figure S1).

As a unique feature of DrugComb, we visualize the synergy scores of a drug combination at each tested dose. The so-called synergy landscape allows a rich information display to facilitate the interpretation of the data, for which the most synergistic and antagonistic doses can be identified separately (Figure 2F). For a given drug or a given cell line, we provide the boxplots and histograms to show the general distributions of the synergy and sensitivity scores, such that the users may assess the general trend. For example, users may evaluate whether drug combinations involving a particular drug tend to be more synergistic, or a cell line tends to be more sensitive to drug treatment. Note that the majority of the data points (93.2%) that we curated from the literature do not contain replicates, and therefore, we decide not to provide the statistical significance of the synergy and sensitivity over a dose-response matrix, as the significance of individual doses contributing to the overall synergy cannot be systematically assessed. Therefore, we would like to highlight the issue of lack of replicates from a typical drug combination screening that may likely hinder the translation of the results into clinical trials.

Network modelling for the mechanisms of action

Once a drug combination experiment has been conducted, for which the results were analysed with the sensitivity and synergy scoring, the next question would be the mechanisms of action of the drug combinations. Network modelling of drug combinations have been recently introduced as an efficient approach for the interpretation of drug combinations, as well as the identification of predictive biomarkers from molecular profiles of cancer (39–42). In DrugComb, the drugs are annotated with their target profiles, and these profiles were further annotated in the signalling networks of cancer cells, such that their first and secondary neighbour proteins can be also retrieved. We utilize the databases including ChEMBL, PubChem and DrugTargetCommons for their primary and secondary targets, and retrieve STITCH for the signalling networks. Furthermore, we have incorporated the transcriptomics profiles of the cancer cell lines into the network, such that their gene expression values can be also displayed (Figure 3A). In addition, we provide the correlation of the gene expression and drug sensitivity such that those neighbouring genes for which their gene expressions are highly correlated with the drug sensitivity will be further identified as potential biomarkers (Figure 3B).

For user-uploaded drug combinations or single drugs, ideally the InChiKeys of the drugs should be provided. This allows the web server to query drug STITCH ID from the major drug databases. In case only the drug names are provided, the web server will query from the major drug databases, for which their targets profiles will be visualized in a generic cancer signalling network. In case the cell line names can be matched with the existing gene expression data, their gene expression values will be displayed as coloured nodes. The network modelling results should be interpreted together with the actual drug screening profiles, such that the drug resistance or sensitivity can be related to its target or neighbouring gene expressions (Figure 3B).

Machine learning for predicting sensitivity and synergy

Upon the large volume of drug combination data curated in DrugComb, we provide the state-of-the-art machine learning algorithms to predict the sensitivity and synergy for a user-selected drug combination on a given cancer cell line.

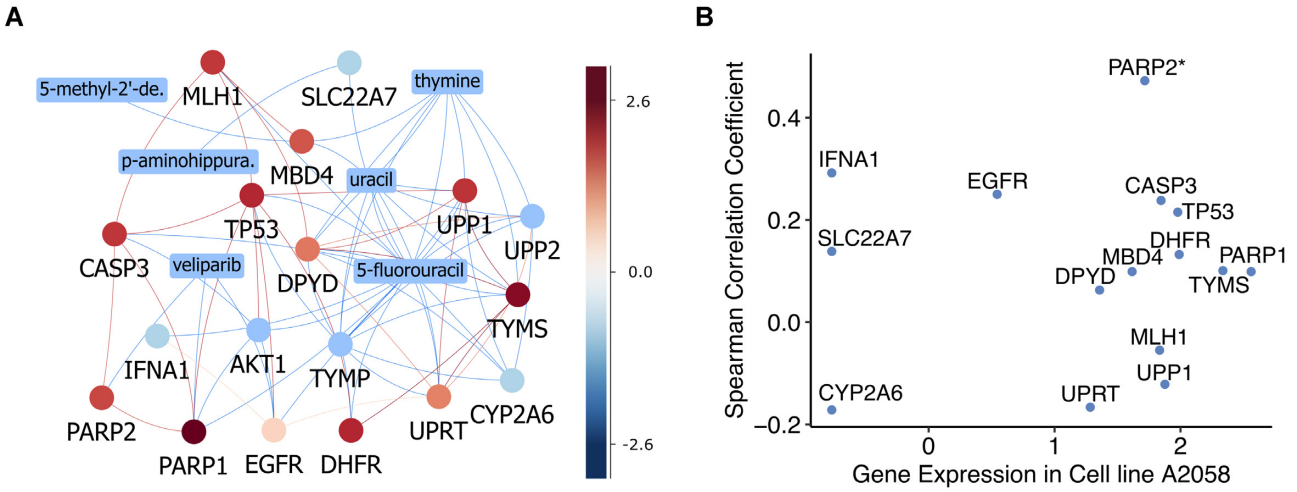


Figure 3. Network modelling of drug combinations. (**A**) An example of veliparib and 5-fluorouracil combination in A2058 melanoma cell line. Drug targets and neighbouring proteins are annotated with their gene expression values, some of which can be modulated by other drugs shown as rectangular boxes. (**B**) Gene expressions of the neighbouring proteins in A2058 as compared with the correlations of these genes with the drug combination sensitivity across all the tested cell lines. PARP2 is the primary target of veliparib, which shows top gene expression in A2058 as well as the highest correlation with the drug combination, suggesting that PARP2 is a potential biomarker for predicting the drug combination sensitivity of veliparib and 5-flurouracil.

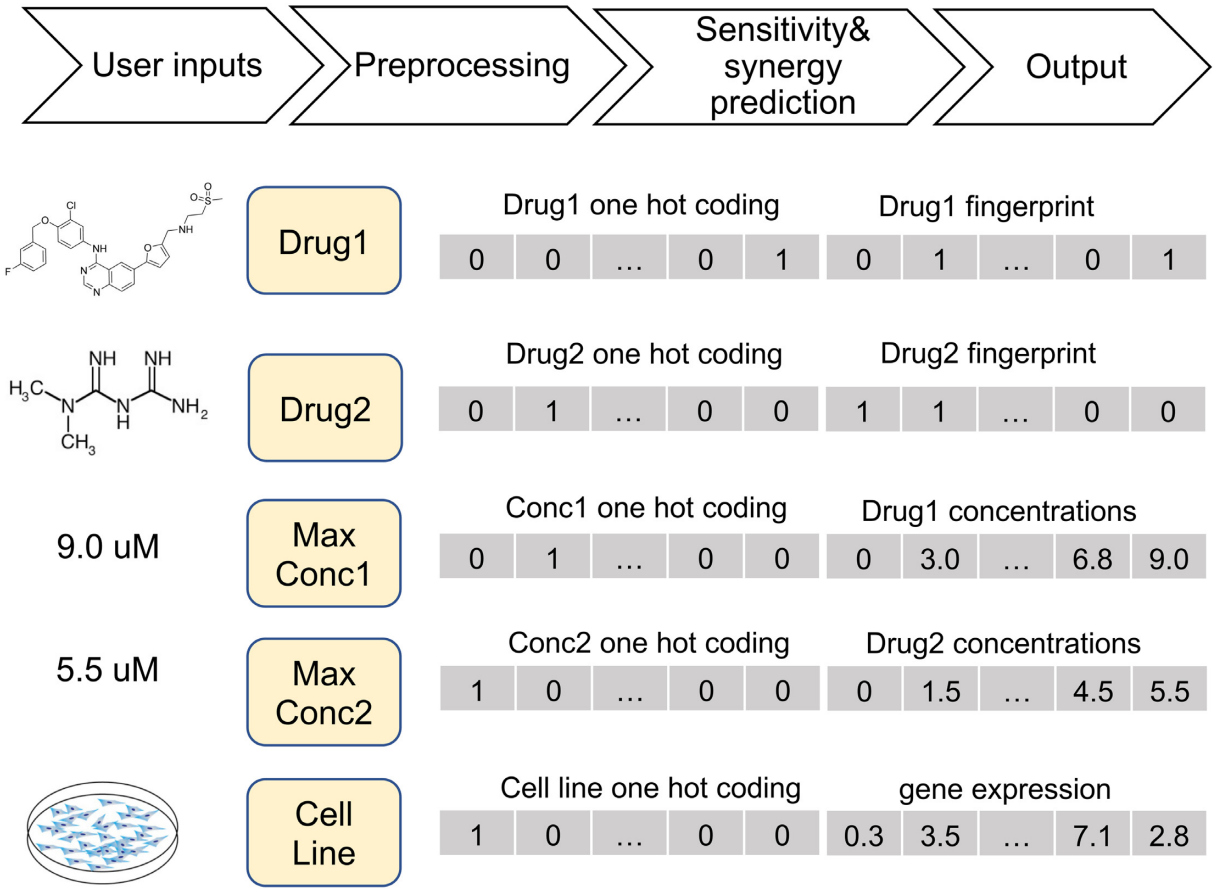


Figure 4. Workflow of machine learning prediction of drug combination sensitivity and synergy. Multiple features are integrated in CatBoost including one-hot encoding of drugs, concentrations and cell lines as well as their specific features including drug chemical fingerprints, drug concentrations and cell line gene expressions.

We utilize the ONEIL data (43) to train a CatBoost model, which has been considered as a reference algorithm for many machine learning tasks (44). The ONEIL data consists of 583 drug combinations involving 38 drugs tested in 39 cell lines, resulting in 92 208 drug combination experiments consisting of 2 305 200 data points. The ONEIL data has been considered a high-quality dataset, as it contains multiple replicates and has been utilized in previous machine learning development (45–47). The CatBoost model is based on decision-trees that can facilitate the integration of different types of features including textual, categorical and numerical values. To build our model, the names of drugs and cell lines are specified as categories in our feature vectors. Additionally, the concentrations for drugs are considered as both numeric values and categories. The cell line's gene expression and compound's structural fingerprints (MACCS) are considered as numerical values. Moreover, in order to accelerate the training process for our model we consider only top 5% most variant genes ($n = 153$) across the 39 cell lines (Figure 4).

Among all the CatBoost hyper-parameters, only four of them show high importance for obtaining the best model. Those hyper-parameters include iterations that indicate the number of trees used in the model, maximum depth of the tree, the learning rate used for gradient steps, and the L2 regularization for the loss function. The best values for mentioned parameters are set and the rest of the hyper-parameters are set to the default values. For drug combination inhibition and synergy scores, a model has been trained separately and the results of the validation accuracy are presented in Table 1.

To facilitate the prediction, users need only to specify the names and the maximal concentrations for each of two drugs, and a cell line name. After receiving the user input, the MACCS fingerprints of the drugs will be obtained by the RCDKlibs package in R, and the cell line gene expression data will be retrieved internally from DrugComb. The pre-processed data will be loaded into the trained models to predict the inhibition values and synergy scores for a 10×10 equally distanced dose matrix within the given maximal concentrations.

Table 1. Prediction accuracy of the CatBoost algorithm tested on ONEIL data

	Correlation	R^2	RMSE
INHIBITION	0.98	0.97	7.12
HSA	0.79	0.62	8.03
ZIP	0.89	0.80	6.55
LOEWE	0.57	0.32	9.68
BLISS	0.87	0.75	7.65

Data contribution

To facilitate the data curation, we have provided a web server for users to upload their drug combination data into the database. The 'Contribute' panel will ask for the annotation information of the drug combination screening results, and then the actual data points will be formulated as a tabular format. We have utilized the contribution module to curate the majority of the literature datasets and found that it greatly facilitates the burden of the data contributors as well as data curators. For example, autofill functions are available when users input the literature citation and drug names. The cell line annotation is also available by retrieving the Cellosaurus website for its disease classification and other cross-reference links. Furthermore, data contributors are guided to provide critical information about assay protocols, such as detection technologies and culture time. When the data has been successfully uploaded, we will first manually check the format, completeness, and validity of the uploaded information, and then integrate them into the database via the data analysis and annotation functions (Figure 5A). In addition to the actual data points as an outcome of such a data curation effort, we can also systematically evaluate the differences in the assay protocols (Figure 5B), which might provide more insights on assessing the reproducibility of the drug sensitivity screens (48). Taken together, we believe that the data contribution may greatly facilitate the open access of drug screening data and therefore we encourage the users of DrugComb to be part of the community-driven data curation team in the future.

Technical aspects

DrugComb is built using PHP 7.4.14 [Laravel Framework 6.20.7] for server-side data processing, Javascript ECMAScript 2015 for the frontend, D3.js 5.7.0, Vis.js 4.18.1 and Plotly library 1.40.0 for the generation of the interactive visualizations. Data is stored in MariaDB 10.3.17 with RMariaDB 1.0.6.9000 as the driver for interfacing with R. Software development tools including Python 3.6.7, numpy 1.14.1, pandas 0.23.4, scikit-learn 0.20.2, RDkit 2018.03.4, R version 3.5.1, synergyfinder 2.2.4 and tidyverse 1.2.1 are used in the analytical pipelines. Linux distribution CentOS-8 with the kernel 4.18.0 64-bit running on four processor cores and 64 Gb of RAM is used for hosting the web service on a computational cluster.

The data portal has been designed in a straightforward manner to maximize the user flexibility to retrieve the existing datasets as well as to analyse their own datasets. We provided the API access at http://api.drugcomb.org such that users can request data as json files. The API is implemented using the PHP laravel framework. Instructions of each of the modules are provided in their associated web pages and the overview of the data portal was summarized as tutorial video available at the home page. We aim to continue accommodating new features such as cloud-based computing and data infrastructure to facilitate the FAIRness (Findable, Accessible, Interoperable and Reusable) of drug screening data analysis. Meanwhile, the community-based features such as data contribution and quality control can be developed further.

DISCUSSION

Making cancer treatment more effective is what a combination therapy aims to achieve. With the advances of high-throughput drug screening technologies, an increasing number of drug combinations have been tested. However, before we can develop robust machine learning and network

A

DrugComb Contribution

test user

Assay Data | Summary Data | Cell line Data | Drug Data

Assay Data

Create a new entry

Show 10 entries

Search:

Action	Assay_type	Plate_format	Cell_per_well	Detection_technology	Negative_control	Negative_contr
Edit Delete	cell viability	384	5000	CellTiter-Glo	DMSO	NA
Edit Delete	cell viability	384	1500	PrestoBlue	DMSO	NA
Edit Delete	cell viability	NA	NA	CellTiter-Glo	DMSO	NA
Edit Delete	caspase	NA	1000	Caspase Glow	DMSO	2000000

Showing 1 to 4 of 4 entries

Previous 1 Next

B

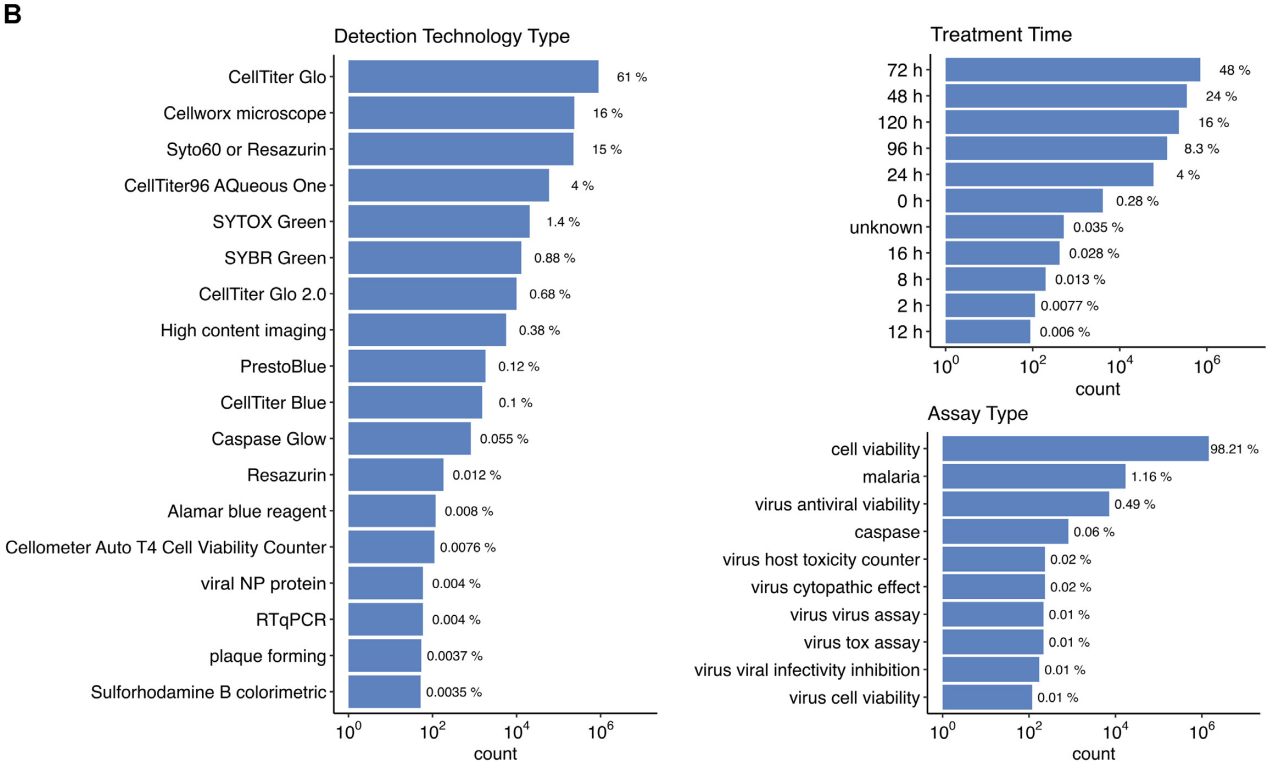


Figure 5. (**A**) The data contribution interface enables a community-driven data curation effort. (**B**) Statistics about assay protocols.

modelling algorithms to predict and understand the potential drug combinations, the datasets need to be systematically curated and harmonized. Here we report the major updates of DrugComb, a comprehensive data portal for the drug discovery community to access the concurrent high-throughput drug combination as well as monotherapy drug screening datasets. These datasets have been deeply curated, standardized and harmonized with the data analysis tools including synergy and sensitivity scoring, such that their potential can be maximized within a unified framework. Furthermore, we have updated the network modelling of the drug combinations, such that the transcriptomics profiles of the cancer cell line, and drug target profiles can be integrated in a signalling network where the protein-protein interactions may provide deeper insights on the mechanisms of drugs and drug combinations. In addition, we have provided a machine learning model to predict a given drug combination for a cell line at the single dose level. To the best of our knowledge, this is the first drug combination prediction tool that has been made online with easy accessibility for drug discovery users.

The four basic modules of DrugComb, i.e. (i) data curation, (ii) synergy and sensitivity scoring, (iii) network modelling and (iv) machine learning constitute a workflow of network pharmacological approaches based on which we may gain deeper understanding of drug-drug interactions. Currently, DrugComb focuses on small molecule drugs such as cytotoxic and kinase inhibitors, while immunotherapy and gene therapy drugs are largely missing. Furture steps of DrugComb will involve constant improvement on the data coverage, for example, by including drugs from other classes. Moreover, we will include higher-order combinations that involve more than two drugs (e.g. (21)). In addition, we will consider the datasets from more recent techniques of microfluidic-based drug screening (49), as well as from patient-derived samples such as 3D organoid-based drug screening (50) and patient-derived xenograft mouse models (51). These datasets may help identify drug combinations that are more translational to the clinics compared to cell line-based studies (9). Meanwhile, the data analysis tools will be also updated to incorporate the new data types. For example, we will develop mathematical and statistical methods for analysing and visualizing higher-order drug combinations. Taken together, we envisage that the high-quality data in DrugComb will serve as a benchmark for the development of more robust and predictive machine learning models, for example, to improve the transfer learning from one study to another study, or to an under-studied tissue (18), as well as accurate network-based models to capture the mechanisms of drug combinations that may eventually lead to predictive biomarkers that warrant patient stratification for maximizing the efficacy of combinatorial therapies.

DATA AVAILABILITY

The synergy and sensitivity scores in DrugComb are freely available for download. Larger batch downloads of raw data are permitted by contacting the authors. The AstraZeneca drug combination datasets are proprietary, and a separate agreement is needed, available at https://openinnovation.astrazeneca.com/. The visualization results for sensitivity, synergy and network models are downloadable as images. The source code for analysing the drug combination datasets is available as the R package SynergyFinder version 2.2.4 (https://bioconductor.org/packages/release/bioc/html/synergyfinder.html). We are committed to open data and welcome any researchers to participate in the development of data curation and harmonization tools for drug discovery.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the authors of the drug combination studies to share their datasets, especially the AstraZeneca for agreeing the Dream Challenge data to be part of DrugComb. We thank also the DepMap consortium and the Cell Model Passports to make the transcriptomics profiles of cancer cell lines freely available. We thank the NCATS and other institutions for making their drug screening datasets easily accessible. The data portal is located at the CSC-IT Center for Science in Finland.

FUNDING

European Research Council (ERC) starting grant DrugComb (Informatics approaches for the rational selection of personalized cancer drug combinations) [716063]; European Commission H2020 EOSC-life (Providing an open collaborative space for digital biology in Europe [824087]; Academy of Finland Research Fellow grant [317680]; Sigrid Jusélius Foundation. Funding for open access charge: ERC starting grant DrugComb. WW and JB hold salaried positions funded by University of Helsinki through Doctoral Program of Biomedicine (DPBM); W.W. also receives a personal grant from K. Albin Johanssons stiftelse.
Conflict of interest statement. None declared.

REFERENCES

1. Campbell,P.J., Getz,G., Korbel,J.O., Stuart,J.M., Jennings,J.L., Stein,L.D., Perry,M.D., Nahal-Bose,H.K., Ouellette,B.F.F., Li,C.H. *et al.* (2020) Pan-cancer analysis of whole genomes. *Nature*, **578**, 82–93.
2. Doroshow,J.H. and Simon,R.M. (2017) On the design of combination cancer therapy. *Cell*, **171**, 1476–1478.
3. He,L., Kulesskiy,E., Saarela,J., Turunen,L., Wennerberg,K., Aittokallio,T. and Tang,J. (2018) Methods for high-throughput drug combination screening and synergy scoring. *Methods Mol. Biol.*, **1711**, 351–398.
4. Lukas,M., Velten,B., Sellner,L., Tomska,K., Hüellein,J., Walther,T., Wagner,L., Muley,C., Wu,B., Oleś,M. *et al.* (2020) Survey of ex vivo drug combination effects in chronic lymphocytic leukemia reveals synergistic drug effects and genetic dependencies. *Leukemia*, **34**, 2934–2950.
5. Tyner,J.W., Tognon,C.E., Bottomly,D., Wilmot,B., Kurtz,S.E., Savage,S.L., Long,N., Schultz,A.R., Traer,E., Abel,M. *et al.* (2018) Functional genomic landscape of acute myeloid leukaemia. *Nature*, **562**, 526–531.
6. Palmer,A.C., Plana,D., Gao,H., Korn,J.M., Yang,G., Green,J., Zhang,X., Velazquez,R., McLaughlin,M.E., Ruddy,D.A. *et al.* (2020) A proof of concept for biomarker-guided targeted therapy against

ovarian cancer based on patient-derived tumor xenografts. *Cancer Res.*, **80**, 4278–4287.
7. Palmer,A.C., Chidley,C. and Sorger,P.K. (2019) A curative combination cancer therapy achieves high fractional cell killing through low cross-resistance and drug additivity. *eLife*, **8**, e50036.
8. Sen,P., Saha,A. and Dixit,N.M. (2019) You cannot have your synergy and efficacy too. *Trends Pharmacol. Sci.*, **40**, 811–817.
9. Palmer,A.C. and Sorger,P.K. (2017) Combination cancer therapy can confer benefit via patient-to-patient variability without drug additivity or synergy. *Cell*, **171**, 1678–1691.
10. Vlot,A.H.C., Aniceto,N., Menden,M.P., Ulrich-Merzenich,G. and Bender,A. (2019) Applying synergy metrics to combination screening data: agreements, disagreements and pitfalls. *Drug Discov. Today*, **24**, 2286–2298.
11. Meyer,C.T., Wooten,D.J., Lopez,C.F. and Quaranta,V. (2020) Charting the fragmented landscape of drug synergy. *Trends Pharmacol. Sci.*, **41**, 266–280.
12. Malyutina,A., Majumder,M.M., Wang,W., Pessia,A., Heckman,C.A. and Tang,J. (2019) Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer. *PLoS Comput. Biol.*, **15**, e1006752.
13. Zagidullin,B., Aldahdooh,J., Zheng,S., Wang,W., Wang,Y., Saad,J., Malyutina,A., Jafari,M., Tanoli,Z., Pessia,A. *et al.* (2019) DrugComb: an integrative cancer drug combination data portal. *Nucleic Acids Res.*, **47**, W43–W51.
14. Zhang,B., Tang,C., Yao,Y., Chen,X., Zhou,C., Wei,Z., Xing,F., Chen,L., Cai,X., Zhang,Z. *et al.* (2021) The tumor therapy landscape of synthetic lethality. *Nat. Commun.*, **12**, 1275.
15. Liu,H., Zhang,W., Zou,B., Wang,J., Deng,Y. and Deng,L. (2020) DrugCombDB: a comprehensive database of drug combinations toward the discovery of combinatorial therapy. *Nucleic Acids Res.*, **48**, D871–D881.
16. Seo,H., Tkachuk,D., Ho,C., Mammoliti,A., Rezaie,A., Madani Tonekaboni,S.A. and Haibe-Kains,B. (2020) SYNERGxDB: an integrative pharmacogenomic portal to identify synergistic drug combinations for precision oncology. *Nucleic Acids Res.*, **48**, W494–W501.
17. Shah,K., Ahmed,M. and Kazi,J.U. (2021) The Aurora kinase/β-catenin axis contributes to dexamethasone resistance in leukemia. *NPJ Precis. Oncol.*, **5**, 13.
18. Kim,Y., Zheng,S., Tang,J., Jim Zheng,W., Li,Z. and Jiang,X. (2021) Anticancer drug synergy prediction in understudied tissues using transfer learning. *J. Am. Med. Informatics Assoc. JAMIA*, **28**, 42–51.
19. Manic,G., Musella,M., Corradi,F., Sistigu,A., Vitale,S., Soliman Abdel Rehim,S., Mattiello,L., Malacaria,E., Galassi,C., Signore,M. *et al.* (2021) Control of replication stress and mitosis in colorectal cancer stem cells through the interplay of PARP1, MRE11 and RAD51. *Cell Death Differ.*, https://doi.org/10.1038/s41418-020-00733-4.
20. Menden,M.P., Wang,D., Mason,M.J., Szalai,B., Bulusu,K.C., Guan,Y., Yu,T., Kang,J., Jeon,M., Wolfinger,R. *et al.* (2019) Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nat. Commun.*, **10**, 2674.
21. Ansbro,M.R., Itkin,Z., Chen,L., Zahoranszky-Kohalmi,G., Amaratunga,C., Miotto,O., Peryea,T., Hobbs,C.V., Suon,S., Sá,J.M. *et al.* (2020) Modulation of triple artemisinin-based combination therapy pharmacodynamics by plasmodium falciparum genotype. *ACS Pharmacol. Transl. Sci.*, **3**, 1144–1157.
22. Bobrowski,T., Chen,L., Eastman,R.T., Itkin,Z., Shinn,P., Chen,C., Guo,H., Zheng,W., Michael,S., Simeonov,A. *et al.* (2021) Synergistic and Antagonistic Drug Combinations against SARS-CoV-2. *Mol. Ther.*, **29**, 873–885.
23. Bairoch,A. (2018) The cellosaurus, a cell-line knowledge resource. *J. Biomol. Tech.*, **29**, 25–38.
24. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B. *et al.* (2021) PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.*, **49**, D1388–D1395.
25. Mendez,D., Gaulton,A., Bento,A.P., Chambers,J., De Veij,M., Félix,E., Magariños,M.P., Mosquera,J.F., Mutowo,P., Nowotka,M. *et al.* (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.*, **47**, D930–D940.
26. Chambers,J., Davies,M., Gaulton,A., Papadatos,G., Hersey,A. and Overington,J.P. (2014) UniChem: extension of InChI-based compound mapping to salt, connectivity and stereochemistry layers. *J. Cheminformatics*, **6**, 43.
27. Wishart,D.S., Feunang,Y.D., Guo,A.C., Lo,E.J., Marcu,A., Grant,J.R., Sajed,T., Johnson,D., Li,C., Sayeeda,Z. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
28. Kanehisa,M., Furumichi,M., Sato,Y., Ishiguro-Watanabe,M. and Tanabe,M. (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.*, **49**, D545–D551.
29. Tang,J., Tanoli,Z.U., Ravikumar,B., Alam,Z., Rebane,A., Vähä-Koskela,M., Peddinti,G., van Adrichem,A.J., Wakkinen,J., Jaiswal,A. *et al.* (2018) Drug target commons: A community effort to build a consensus knowledge base for drug-target interactions. *Cell Chem. Biol.*, **25**, 224–229.
30. Szklarczyk,D., Santos,A., von Mering,C., Jensen,L.J., Bork,P. and Kuhn,M. (2016) STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.*, **44**, D380–D384.
31. UniProt,C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
32. Tsherniak,A., Vazquez,F., Montgomery,P.G., Weir,B.A., Kryukov,G., Cowley,G.S., Gill,S., Harrington,W.F., Pantel,S., Krill-Burger,J.M. *et al.* (2017) Defining a cancer dependency map. *Cell*, **170**, 564–576.
33. van der Meer,D., Barthorpe,S., Yang,W., Lightfoot,H., Hall,C., Gilbert,J., Francies,H.E. and Garnett,M.J. (2019) Cell model Passports-a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res.*, **47**, D923–D929.
34. Visser,U., Abeyruwan,S., Vempati,U., Smith,R.P., Lemmon,V. and Schürer,S.C. (2011) BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results. *BMC Bioinformatics*, **12**, 257.
35. Douglass,E.F., Allaway,R.J., Szalai,B., Wang,W., Tian,T., Fernández-Torras,A., Realubit,R., Karan,C., Zheng,S., Pessia,A. *et al.* (2020) A community challenge for pancancer drug mechanism of action inference from perturbational profile data. biorXiv doi: https://doi.org/10.1101/2020.12.21.423514, 23 December 2020, preprint: not peer reviewed.
36. Yadav,B., Wennerberg,K., Aittokallio,T. and Tang,J. (2015) Searching for drug synergy in complex dose-response landscapes using an interaction potency model. *Comput. Struct. Biotechnol. J.*, **13**, 504–513.
37. Tang,J., Wennerberg,K. and Aittokallio,T. (2015) What is synergy? The Saariselkä agreement revisited. *Front. Pharmacol.*, **6**, 181.
38. Walker,T., Mitchell,C., Park,M.A., Yacoub,A., Graf,M., Rahmani,M., Houghton,P.J., Voelkel-Johnson,C., Grant,S. and Dent,P. (2009) Sorafenib and vorinostat kill colon cancer cells by CD95-dependent and -independent mechanisms. *Mol. Pharmacol.*, **76**, 342–355.
39. Cheng,F., Kovács,I.A. and Barabási,A.L. (2019) Network-based prediction of drug combinations. *Nat. Commun.*, **10**, 1197.
40. Zhou,Y., Hou,Y., Shen,J., Huang,Y., Martin,W. and Cheng,F. (2020) Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.*, **6**, 14.
41. Tang,J., Gautam,P., Gupta,A., He,L., Timonen,S., Akimov,Y., Wang,W., Szwajda,A., Jaiswal,A., Turei,D. *et al.* (2019) Network pharmacology modeling identifies synergistic Aurora B and ZAK interaction in triple-negative breast cancer. *NPJ Syst. Biol. Applic.*, **5**, 20.
42. Liu,Q. and Xie,L. (2021) TranSynergy: mechanism-driven interpretable deep neural network for the synergistic prediction and pathway deconvolution of drug combinations. *PLoS Comput. Biol.*, **17**, e1008653.
43. O'Neil,J., Benita,Y., Feldman,I., Chenard,M., Roberts,B., Liu,Y., Li,J., Kral,A., Lejnine,S., Loboda,A. *et al.* (2016) An unbiased oncology compound screen to identify novel combination strategies. *Mol. Cancer Ther.*, **15**, 1155–1162.
44. Hancock,J.T. and Khoshgoftaar,T.M. (2020) CatBoost for big data: an interdisciplinary review. *J. Big Data*, **7**, 94.
45. Preuer,K., Lewis,R.P.I., Hochreiter,S., Bender,A., Bulusu,K.C. and Klambauer,G. (2018) DeepSynergy: predicting anti-cancer drug synergy with Deep Learning. *Bioinformatics*, **34**, 1538–1546.

46. Ling,A. and Huang,R.S. (2020) Computationally predicting clinical drug combination efficacy with cancer cell line screens and independent drug action. *Nat. Commun.*, **11**, 5848.
47. Julkunen,H., Cichonska,A., Gautam,P., Szedmak,S., Douat,J., Pahikkala,T., Aittokallio,T. and Rousu,J. (2020) Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nat. Commun.*, **11**, 6136.
48. Haverty,P.M., Lin,E., Tan,J., Yu,Y., Lam,B., Lianoglou,S., Neve,R.M., Martin,S., Settleman,J., Yauch,R.L. *et al.* (2016) Reproducible pharmacogenomic profiling of cancer cell line panels. *Nature*, **533**, 333–337.
49. Eduati,F., Utharala,R., Madhavan,D., Neumann,U.P., Longerich,T., Cramer,T., Saez-Rodriguez,J. and Merten,C.A. (2018) A microfluidics platform for combinatorial drug screening on cancer biopsies. *Nat. Commun.*, **9**, 2434.
50. Du,Y., Li,X., Niu,Q., Mo,X., Qui,M., Ma,T., Kuo,C.J. and Fu,H. (2020) Development of a miniaturized 3D organoid culture platform for ultra-high-throughput screening. *J. Mol. Cell Biol.*, **12**, 630–643.
51. Gao,H., Korn,J.M., Ferretti,S., Monahan,J.E., Wang,Y., Singh,M., Zhang,C., Schnell,C., Yang,G., Zhang,Y. *et al.* (2015) High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. *Nat. Med.*, **21**, 1318–1325.